Report – Week 5 – Capstone project

Target of this Project:
New York City Food venue Cluster locators, Popular food venues, Most populated food venues, popularity of various cuisines

1. Introduction where you discuss the business problem and who would be interested in this project.

Scope of the project

- **Designated clients** would be the one who are interested to use this quantifiable analysis to understand the distribution of different cultures and cuisines over "the most diverse city on the planet - NYC".
- Also, this project can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.
- Finding the best locality for a targeted cuisine.

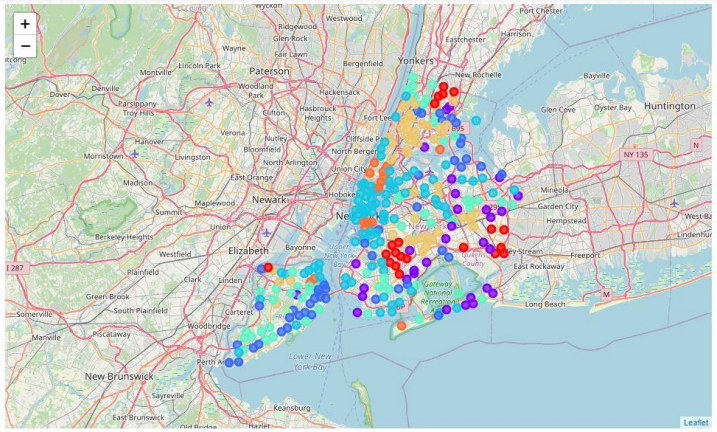
2. Data where you describe the data that will be used to solve the problem and the source of the data.

- Collect the new york city data from ('https://cocl.us/new_york_dataset/newyork_data.json')
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are unique Restaurants.
- Using rating for each resturant , we will sort that data.
- Visualize the Ranking of neighbourhoods using folium library(python)

3.Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

1. This project will help to understand the diversity of a neighbourhood by leveraging venue data from Four square's 'Places API' and 'k-means clustering' machine learning algorithm.
2. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighbourhood.
3. **Clients** would be the one who are interested to use this quantifiable analysis to understand the distribution of different cultures and cuisines over "the most diverse city on the planet - NYC".
4. Also, this project can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.
5. Visualization charts like stiletto method and bar charts

5. Results section where you discuss the results.



In [72]:
```
# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = nyc_grouped['Neighborhood']
```

In [73]:
```
for ind in np.arange(nyc_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(nyc_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

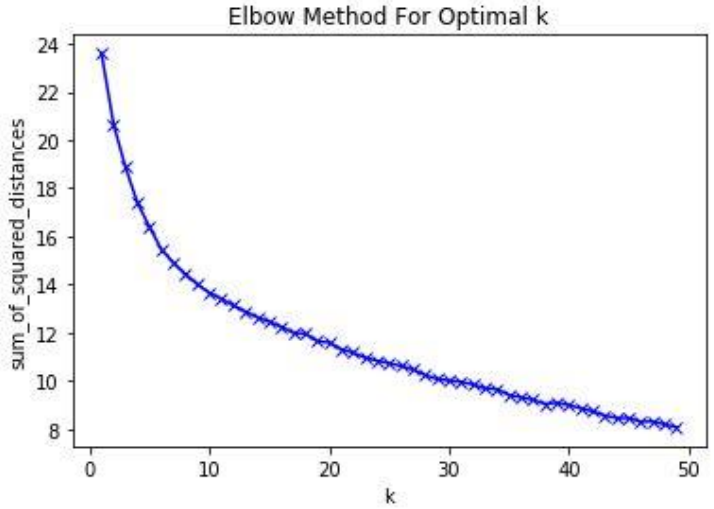
Out[73]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allerton	Pizza Place	Chinese Restaurant	Mexican Restaurant	Fried Chicken Joint	Fast Food Restaurant
1	Annadale	Pizza Place	Sushi Restaurant	American Restaurant	Italian Restaurant	Middle Eastern Restaurant
2	Arden Heights	Pizza Place	American Restaurant	Italian Restaurant	Sushi Restaurant	Mexican Restaurant
3	Arlington	Pizza Place	Spanish Restaurant	American Restaurant	Fast Food Restaurant	Polish Restaurant
4	Arrochar	Italian Restaurant	Pizza Place	Japanese Restaurant	Polish Restaurant	Latin American Restaurant

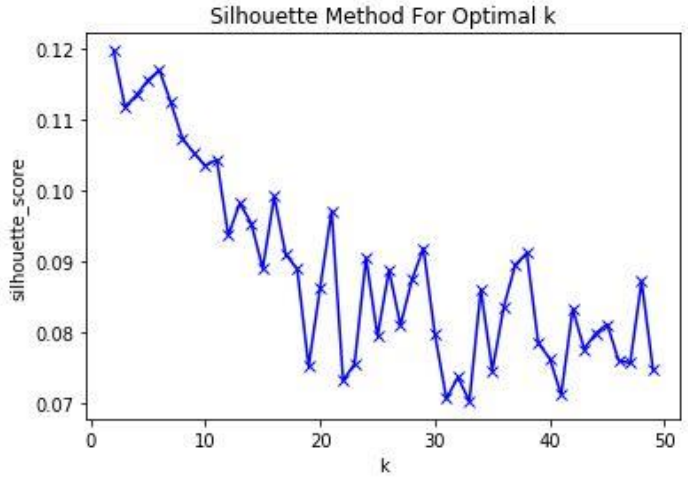
```
sum_of_squared_distances = []
K = range(1,50)
for k in K:
    print(k, end=' ')
    kmeans = KMeans(n_clusters=k).fit(nyc_grouped_clustering)
    sum_of_squared_distances.append(kmeans.inertia_)
```

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 4
6 47 48 49
```

```
plt.plot(K, sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('sum_of_squared_distances')
plt.title('Elbow Method For Optimal k');
```



In [78]:
```
plt.plot(K_sil, sil, 'bx-')
plt.xlabel('k')
plt.ylabel('silhouette_score')
plt.title('Silhouette Method For Optimal k')
plt.show()
```



There is a peak at k = 2, k = 4 and k = 8. Two and four clusters will give a very broad classification of the venues.

6. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results

```
In [98]: for col in required_column:
             print(cluster_6[col].value_counts(ascending = Fal
             print("-----------------------------------------
```

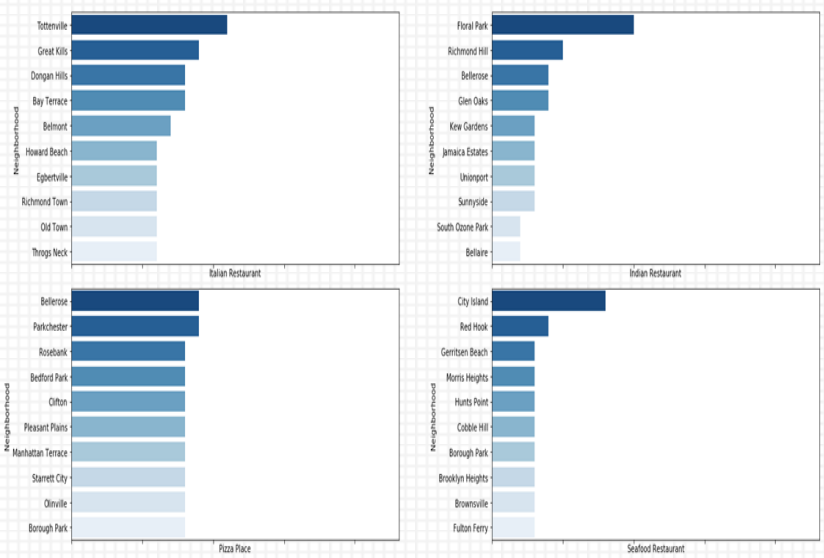
```
Pizza Place             6
Fast Food Restaurant    3
Indian Restaurant       2
Mexican Restaurant      1
Thai Restaurant         1
Name: 1st Most Common Venue, dtype: int64
-----------------------------------------
Pizza Place                  3
Fast Food Restaurant         2
Chinese Restaurant           2
Middle Eastern Restaurant    1
Seafood Restaurant           1
Thai Restaurant              1
Spanish Restaurant           1
Italian Restaurant           1
Mexican Restaurant           1
Name: 2nd Most Common Venue, dtype: int64
-----------------------------------------
Queens      6
Brooklyn    6
Bronx       1
Name: Borough, dtype: int64
-----------------------------------------
```

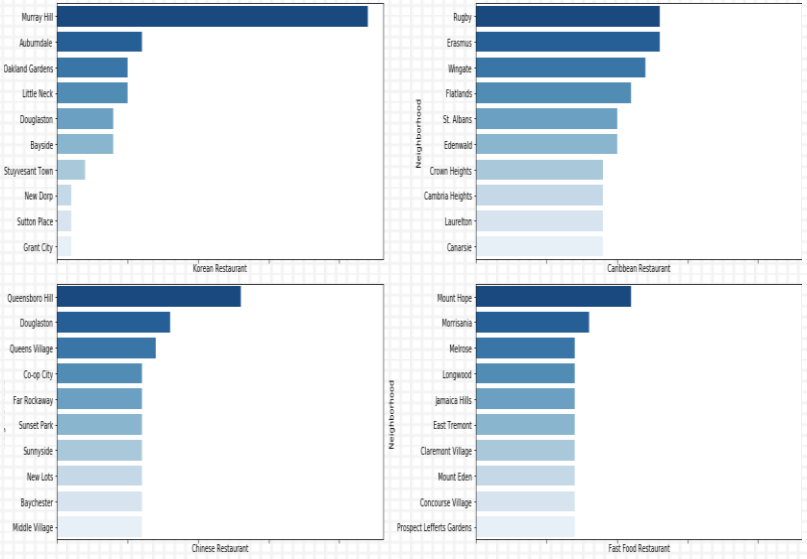
The top restuarants are

1. Deli / Bodega 1253
2. Pizza Place 1082
3. Coffee Shop 941
4. Chinese Restaurant 678
5. Donut Shop 643
6. Fast Food Restaurant 598
7. Bakery 589
8. Italian Restaurant 446
9. Bagel Shop 404
10. Café 383
11. Mexican Restaurant 373

"Based on observations an Italian restaurant in Borough neighbourhood will be a good investment"
Further more, any pizza place or fast food new set up wouldn't be recommended as the area has high density of the same.

7.Conclusion section where you conclude the report.






```
Pizza Place          1
Greek Restaurant     1
Name: 1st Most Common Venue, dtype: int64
----------------------------------------------
Sushi Restaurant     1
Pizza Place          1
Name: 2nd Most Common Venue, dtype: int64
----------------------------------------------
Staten Island     1
Queens            1
Name: Borough, dtype: int64
----------------------------------------------
```

Conclusion:
Greek food and pizzerias are very common in NYC
Korean and chinese too.
Introduces more pizzerias have lesser chances of profits as its plentiful
Lebanese and Arabic food is less. Probably something with rice and kebobs, they could work in nyc.
Perhaps a bagel shop with fresh seafood options like bagel with tuna, since nyc is geographically benefiting from its shore position and cafes are popular in NYC.

Thank you for participating in this course and journey